



11/No fee
A/Box Sequence

GTN 901
187487032

NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO *HELICOBACTER*
PYLORI FOR DIAGNOSTICS AND THERAPEUTICS

Background of the Invention

Helicobacter pylori is a gram-negative, S-shaped, microaerophilic bacterium that was discovered and cultured from a human gastric biopsy specimen. (Warren, J.R. and B. Marshall, (1983) *Lancet* 1: 1273-1275; and Marshall et al., (1984) *Microbios Lett.* 25: 83-88). *H. pylori* has been strongly linked to chronic gastritis and duodenal ulcer disease. (Rathbone et. al., (1986) *Gut* 27: 635-641). Moreover, evidence is accumulating for an etiologic role of *H. pylori* in nonulcer dyspepsia, gastric ulcer disease, and gastric adenocarcinoma. (Blaser M. J., (1993) *Trends Microbiol.* 1: 255-260). Transmission of the bacteria occurs via the oral route, and the risk of infection increases with age. (Taylor, D.N. and M. J. Blaser, (1991) *Epidemiol. Rev* 13: 42-50). *H. pylori* colonizes the human gastric mucosa, establishing an infection that usually persists for decades. Infection by *H. pylori* is prevalent worldwide. Developed countries have infection rates over 50% of the adult population, while developing countries have infection rates reaching 90% of the adults over the age of 20. (Hopkins R. J. and J. G. Morris (1994) *Am. J. Med.* 97: 265-277).

The bacterial factors necessary for colonization of the gastric environment, and for virulence of this pathogen, are poorly understood. Examples of the putative virulence factors include the following: urease, an enzyme that may play a role in neutralizing gastric acid pH (Eaton et al., (1991) *Infect. Immunol.* 59: 2470-2475; Ferrero, R.L. and A. Lee (1991) *Microb. Ecol. Hlth. Dis.* 4: 121-134; Labigne et al., (1991) *J. Bacteriol.* 173: 1920-1931); the bacterial flagellar proteins responsible for motility across the mucous layer. Hazell et al., (1986) *J. Inf. Dis.* 153: 658-663; Leying et al., (1992) *Mol. Microbiol.* 6: 2863-2874; and Haas et al., (1993) *Mol. Microbiol.* 8: 753-760; Vac A, a bacterial toxin that induces the formation of intracellular vacuoles in epithelial cells Schmitt, W. and R. Haas, (1994) *Molecular Microbiol.* 12(2): 307-319; and several gastric tissue-specific adhesins. Boren et al., (1993) *Science* 262: 1892-1895; Evans et al., (1993) *J. Bacteriol.* 175: 674-683; and Falk et al., (1993) *Proc. Natl. Acad. Sci. USA* 90: 2035-2039.

Numerous therapeutic agents are currently available that eradicate *H. pylori* infections *in vitro*. (Huesca et. al., (1993) *Zbl. Bakt.* 280: 244-252; Hopkins, R. J. and J. G. Morris, *supra*). However, many of these treatments are suboptimally effective *in vivo* because of bacterial resistance, altered drug distribution, patient non-compliance or poor drug availability. (Hopkins, R. J. and J. G. Morris, *supra*). Treatment with antibiotics combined with bismuth are part of the standard regime used to treat *H. pylori* infection. (Malfertheiner, P. and J. E. Dominguez-Munoz (1993) *Clinical Therapeutics* 15 Supp. B: 37-48). Recently, combinations of a proton pump inhibitors and a single antibiotic have been shown to

ameliorate duodenal ulcer disease. (Malfertheiner, P. and J. E. Dominguez-Munoz *supra*). However, methods employing antibiotic agents can have the problem of the emergence of bacterial strains which are resistant to these agents. (Hopkins, R. J. and J. G. Morris, *supra*). These limitations demonstrate that new more effective methods are needed to combat *H. pylori* infections *in vivo*. In particular, the design of new vaccines that may prevent infection by this bacterium is highly desirable.

Summary of the Invention

This invention relates to novel genes, e.g., genes encoding bacterial surface proteins, from the organism *Helicobacter pylori*, and other related genes, their products, and uses thereof. The nucleic acids and peptides of the present invention have utility for diagnostic and therapeutics for *H. pylori* and other *Helicobacter* species. They can also be used to detect the presence of *H. pylori* and other *Helicobacter* species in a sample; and for use in screening compounds for the ability to interfere with the *H. pylori* life cycle or to inhibit *H. pylori* infection. More specifically, this invention features compositions of nucleic acids corresponding to entire coding sequences of *H. pylori* surface proteins or parts thereof, nucleic acids capable of binding mRNA from *H. pylori* surface proteins to block protein translation, and methods for producing *H. pylori* surface proteins or parts thereof using peptide synthesis and recombinant DNA techniques. This invention also includes antibodies and nucleic acids sequences useful as probes to detect *H. pylori* infection. In addition, vaccine compositions and methods for the protection against infection by *H. pylori* are described.

Brief Description of the Drawings

Figures 1A - 558A contain the amino acid sequences of the *H. pylori* polypeptides of this invention (HPP1A - HPP558A).

Figure 559 contains nucleic acid sequences which encode *H. pylori* polypeptides of this invention.

Figure 560 is a table which contains information from homology searches performed on the sequences of this invention using the BLAST algorithm contained in the Wisconsin Sequence Analysis Package.

Detailed Description of the Invention

Definitions

A purified preparation or a substantially pure preparation of a polypeptide, as used herein, means a polypeptide that has been separated from other proteins, lipids, and nucleic acids with which it naturally occurs. Preferably, the polypeptide is also separated from

substances, e.g., antibodies or gel matrix, e.g., polyacrylamide, which are used to purify it. Preferably, the polypeptide constitutes at least 10, 20, 50 70, 80 or 95% dry weight of the purified preparation. Preferably, the preparation contains: sufficient polypeptide to allow protein sequencing; at least 1, 10, or 100 μ g of the polypeptide; at least 1, 10, or 100 mg of the polypeptide.

A purified preparation of cells refers to, in the case of plant or animal cells, an in vitro preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

The metabolism of a substance, as used herein, means any aspect of the, expression, function, action, or regulation of the substance. The metabolism of a substance includes modifications, e.g., covalent or non covalent modifications of the substance. The metabolism of a substance includes modifications, e.g., covalent or non covalent modification, the substance induces in other substances. The metabolism of a substance also includes changes in the distribution of the substance. The metabolism of a substance includes changes the substance induces in the distribution of other substances.

A substantially pure nucleic acid, e.g., a substantially pure DNA, is a nucleic acid which is one or both of: not immediately contiguous with both of the coding sequences with which it is immediately contiguous (i.e., one at the 5' end and one at the 3' end) in the naturally-occurring genome of the organism from which the nucleic acid is derived; or which is substantially free of a nucleic acid sequence with which it occurs in the organism from which the nucleic acid is derived. The term includes, for example, a recombinant DNA which is incorporated into a vector, e.g., into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other DNA sequences. Substantially pure DNA also includes a recombinant DNA which is part of a hybrid gene encoding additional *H. pylori* DNA sequence.

Homologous refers to the sequence similarity between two polypeptide molecules or between two nucleic acid molecules. When a position in both of the two compared sequences is occupied by the same base or amino acid monomer subunit, e.g., if a position in each of two DNA molecules is occupied by adenine, then the molecules are homologous at that position. The percent of homology between two sequences is a function of the number of matching or homologous positions shared by the two sequences divided by the number of positions compared x 100. For example, if 6 of 10, of the positions in two sequences are matched or homologous then the two sequences are 60% homologous. By way of example, the DNA sequences ATTGCC and TATGGC share 50% homology. Generally, a comparison is made when two sequences are aligned to give maximum homology.

The terms peptides, proteins, and polypeptides are used interchangeably herein.

As used herein, the term "surface protein" refers to all surface accessible proteins, e.g., inner and outer membrane proteins, proteins adhering to the cell wall, and secreted proteins.

As used herein, the term "transgene" means a nucleic acid sequence (encoding, e.g., one or more polypeptides), which is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the cell's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of the selected nucleic acid, all operably linked to the selected nucleic acid, and may include an enhancer sequence.

As used herein, the term "transgenic cell" refers to a cell containing a transgene.

As used herein, a "transgenic animal" is any animal in which one or more, and preferably essentially all, of the cells of the animal includes a transgene. The transgene can be introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA.

As used herein, the term "cell-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue. The term also covers so-called "leaky" promoters, which regulate expression of a selected DNA primarily in one tissue, but cause expression in other tissues as well.

A polypeptide has *H. pylori* biological activity if it has one, two, three, and preferably more of the following properties: (1) if when expressed in the course of an *H. pylori* infection, it can promote, or mediate the attachment of *H. pylori* to a cell (2) it has an enzymatic activity characteristic of an *H. pylori* protein (3) or the gene which encodes it can rescue a lethal mutation in an *H. pylori* gene. A polypeptide has biological activity if it is an antagonist, agonist, or super-agonist of a polypeptide having one of the above-listed properties.

Misexpression, as used herein, refers to a non-wild type pattern of gene expression. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing

size, amino acid sequence, post-translational modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

As used herein, "host cells" and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refers to cells which can become or have been used as recipients for a recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood by individuals skilled in the art that the progeny of a single parental cell may not necessarily be completely identical in genomic or total DNA compliment to the original parent, due to accident or deliberate mutation.

As used herein, the term "control sequence" refers to a nucleic acid having a base sequence which is recognized by the host organism to effect the expression of encoded sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include a promoter, ribosomal binding site and terminators; in eukaryotes, generally such control sequences include promoters, terminators and in some instances, enhancers. The term control sequence is intended to include at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences.

As used herein, the term "operably linked" refers to sequences joined or ligated to function in their intended manner. For example, a control sequence is operably linked to coding sequence by ligation in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequence and host cell.

An "open reading frame", also referred to herein as ORF, is a region of nucleic acid which encodes a peptide. This region may represent a portion of a coding sequence or a total sequence.

As used herein, a "coding sequence" is a nucleic acid sequence which is transcribed into messenger RNA and/or translated into a peptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the five prime terminus and a translation stop code at the three prime terminus. A coding sequence can include but is not limited to messenger RNA, synthetic DNA, and recombinant nucleic acid sequences.

A "gene product" is a protein or structural RNA which is specifically encoded for by a gene.

As used herein, the term "probe" refers to a nucleic acid, peptide or other chemical entity which specifically binds to a molecule of interest. Probes are often associated with or capable of associating with a label. A label is a chemical moiety capable of detection. Typical

labels comprise dyes, radioisotopes, luminescent and chemiluminescent moieties, fluorophores, enzymes, precipitating agents, amplification sequences, and the like. Similarly, a nucleic acid, peptide or other chemical entity which specifically binds to a molecule of interest and immobilizes such molecule is referred herein as a "capture ligand". Capture ligands are typically associated with or capable of associating with a support such as nitro-cellulose, glass, nylon membranes, beads, particles and the like. The specificity of hybridization is dependent on conditions such as the base pair composition of the nucleotides, and the temperature and salt concentration of the reaction. These conditions are readily discernable to one of ordinary skill in the art using routine experimentation.

The experimental manipulation of such conditions has been well described in the literature including such books as *Molecular Cloning: A Laboratory Manual*, Sambrook, J., Fritsch, E.F., Maniatis, T., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2nd ed. (1989).

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See e.g., Sambrook, Fritsch, and Maniatis, *Molecular Cloning: Laboratory Manual* 2nd ed. (1989); *DNA Cloning*, Volumes I and II (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); the series, *Methods in Enzymology* (Academic Press, Inc.), particularly Vol. 154 and Vol. 155 (Wu and Grossman, eds.) and *PCR-A Practical Approach* (McPherson, Quirke, and Taylor, eds., 1991).

In one aspect, the invention features a recombinant or substantially pure preparation of HPP1 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP1 polypeptide. The HPP polypeptide sequences described herein are contained in Figures 1A-558A, and the nucleic acid sequences encoding HPP polypeptides are contained in Figure 559. Figure 559 contains nucleic acid sequences which encode polypeptides contained in Figures 1A-558A. The SEQ ID NO's for the polypeptides are set forth in Figures 1A-558A. The nucleic acid sequence (if present in Figure 559) encoding a particular polypeptide has a SEQ ID NO which corresponds to or is similar to the SEQ ID NO of its encoded polypeptide except for the terminal "-AA" portion of the SEQ ID NO of the polypeptide is replaced with ".nt" when referring to a nucleic acid sequence.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP2 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP2 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP3 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP3 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP4 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP4 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP5 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP5 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP6 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP6 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP7 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP7 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP8 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP8 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP9 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP9 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP10 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP10 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP11 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP11 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP12 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP12 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP13 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP13 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP14 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP14 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP15 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP15 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP16 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP16 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP17 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP17 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP18 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP18 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP19 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP19 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP20 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP20 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP21 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP21 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP22 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP22 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP23 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP23 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP24 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP24 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP25 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP25 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP26 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP26 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP27 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP27 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP28 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP28 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP29 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP29 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP30 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP30 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP31 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP31 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP32 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP32 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP33 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP33 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP34 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP34 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP35 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP35 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP36 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP36 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP37 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP37 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP38 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP38 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP39 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP39 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP40 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP40 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP41 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP41 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP42 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP42 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP95 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP95 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP96 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP96 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP97 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP97 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP98 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP98 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP99 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP99 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP100 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP100 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP101 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP101 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP102 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP102 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP103 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP103 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP104 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP104 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP105 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP105 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP106 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP106 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP107 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP107 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP147 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP147 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP148 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP148 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP149 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP149 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP150 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP150 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP151 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP151 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP152 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP152 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP153 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP153 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP154 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP154 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP155 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP155 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP156 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP516 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP157 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP157 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP158 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP158 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP159 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP159 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP225 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP25 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP226 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP226 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP227 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP227 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP228 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP228 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP229 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP229 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP230 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP230 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP231 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP231 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP232 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP232 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP233 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP233 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP234 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP234 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP235 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP235 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP236 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP236 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP237 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP237 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP238 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP238 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP239 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP239 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP240 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP240 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP241 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP241 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP242 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP242 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP243 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP243 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP244 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP244 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP245 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP45 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP246 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP246 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP247 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP247 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP248 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP248 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP249 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP249 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP250 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP250 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP251 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP251 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP52 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP52 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP53 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP53 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP54 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP54 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP255 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP255 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP256 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP256 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP257 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP257 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP258 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP258 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP259 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP259 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP260 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP260 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP261 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP261 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP262 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP262 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP263 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP263 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP290 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP290 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP291 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP291 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP292 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP292 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP293 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP293 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP294 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP294 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP295 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP295 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP296 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP296 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP297 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP297 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP298 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP298 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP299 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP299 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP300 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP300 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP301 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP301 poly peptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP302 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP302 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP329 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP329 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP330 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP330 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP331 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP331 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP332 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP332 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP333 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP333 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP34 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP34 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP35 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP35 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP36 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP36 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP37 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP37 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP38 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP38 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP39 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP39 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP340 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP340 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP341 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP341 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP394 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP394 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP395 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP395 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP396 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP396 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP397 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP397 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP398 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP398 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP399 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP399 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP400 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP400 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP401 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP401 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP402 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP402 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP403 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP403 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP404 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP404 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP405 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP405 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP406 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP406 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP459 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP59 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP460 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP460 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP461 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP461 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP462 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP462 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP463 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP463 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP464 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP464 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP465 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP465 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP466 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP466 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP467 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP467 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP468 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP468 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP469 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP469 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP470 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP470 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP471 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP471 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP498 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP498 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP499 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP499 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP500 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP500 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP501 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP501 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP502 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP502 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP503 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP503 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP504 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP504 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP505 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP505 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP506 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP506 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP507 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP507 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP508 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP508 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP509 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP509 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP510 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP510 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP511 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP511 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP512 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP512 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP513 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP513 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP514 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP514 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP515 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP515 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP516 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP516 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP517 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP517 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP518 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP518 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP519 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP519 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP520 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP520 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP521 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP521 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP522 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP22 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP523 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP523 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP550 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP550 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP551 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP551 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP552 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP552 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP553 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP553 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP554 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP554 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP555 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP555 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP556 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP556 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP557 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP557 polypeptide.

In another aspect, the invention features a recombinant or substantially pure preparation of an HPP558 polypeptide. The invention also includes substantially pure nucleic acid encoding an HPP558 polypeptide.

In another aspect, the invention comprises nucleic acids capable of binding mRNA of *H. pylori*. Such nucleic acid is capable of acting as anti-sense nucleic acid to control the translation of mRNA of *H. pylori*. A further aspect features a nucleic acid which is capable of binding specifically to an *H. pylori* nucleic acid. Such nucleic acid has utility as probes and as capture reagents.

In another aspect, the invention features an expression system comprising an open reading frame corresponding to *H. pylori* nucleic acid. The nucleic acid further comprises a control sequence compatible with an intended host. The expression system is useful for making polypeptides corresponding to *H. pylori* nucleic acid.

In another aspect, the invention features a cell transformed with the expression system to make *H. pylori* polypeptides.

In another aspect, the invention features a method of generating antibodies against *H. pylori* proteins which are capable of binding specifically to *H. pylori* proteins. Such antibody has utility as reagents for immunoassays to evaluate the abundance and distribution of *H. pylori*-specific antigens.

In another aspect, the invention features a method of generating vaccines for immunizing an individual against *H. pylori*. The method includes: immunizing a subject with an *H. pylori* protein, e.g., a surface protein, or portion thereof, and a pharmaceutically acceptable carrier. Such vaccines have therapeutic and prophylactic utilities.

In another aspect, the invention provides a method for generating a vaccine comprising a modified immunogenic *H. pylori* protein, e.g., a surface protein, or portion thereof, and a pharmacologically acceptable carrier.

In another aspect, the invention features a method of evaluating a compound, e.g. a polypeptide, e.g., a fragment of a host cell polypeptide, for the ability to bind an *H. pylori* polypeptide. The method includes: contacting the candidate compound with an *H. pylori* polypeptides and determining if the compound binds or otherwise interacts with an *H. pylori* polypeptide. Compounds which bind *H. pylori* are candidates as activators or inhibitors of the bacterial life cycle. These assays can be performed *in vitro* or *in vivo*.

In another aspect, the invention features a method of evaluating a compound, e.g. a polypeptide, e.g., a fragment of a host cell polypeptide, for the ability to bind an *H. pylori* nucleic acid, e.g., DNA or RNA. The method includes: contacting the candidate compound with an *H. pylori* polypeptides and determining if the compound binds or otherwise interacts with an *H. pylori* polypeptide. Compounds which bind *H. pylori* are candidates as activators or inhibitors of the bacterial life cycle. These assays can be performed *in vitro* or *in vivo*.

The invention features, *H. pylori* polypeptides, preferably a substantially pure preparation of an *H. pylori* polypeptide, or a recombinant *H. pylori* polypeptide. In preferred embodiments: the polypeptide has biological activity; the polypeptide has an amino acid sequence at least 60%, 80%, 90%, 95%, 98%, or 99% homologous to an amino acid sequence contained in Figures 1A-558A; the polypeptide has an amino acid sequence essentially the same as an amino acid sequence in Figures 1A-558A; the polypeptide is at least 5, 10, 20, 50, 100, or 150 amino acids in length; the polypeptide includes at least 5, preferably at least 10, more preferably at least 20, more preferably at least 50, 100, or 150 contiguous amino acids contained in Figures 1A-558A.

In preferred embodiments: the *H. pylori* polypeptide is encoded by a nucleic acid in Figure 559, or by a nucleic acid having at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% homology with a nucleic acid shown in Figure 559.

In a preferred embodiment, the subject *H. pylori* polypeptide differs in amino acid sequence at 1, 2, 3, 5, 10 or more residues from a sequence in Figures 1A-558A. The differences, however, are such that: the *H. pylori* polypeptide exhibits an *H. pylori* biological

activity, e.g., the *H. pylori* polypeptide retains a biological activity of a naturally occurring *H. pylori* enzyme.

In preferred embodiments, the polypeptide includes all or a fragment of an amino acid sequence contained in Figures 1A-558A; fused, in reading frame, to additional amino acid residues, preferably to residues encoded by genomic DNA 5' to the genomic DNA which encodes a sequence contained in Figures 1A-558A.

In yet other preferred embodiments, the *H. pylori* polypeptide is a recombinant fusion protein having a first *H. pylori* polypeptide portion and a second polypeptide portion, e.g., a second polypeptide portion having an amino acid sequence unrelated to *H. pylori*. The second polypeptide portion can be, e.g., any of glutathione-S-transferase, a DNA binding domain, or a polymerase activating domain. In preferred embodiment the fusion protein can be used in a two-hybrid assay.

Polypeptides of the invention include those which arise as a result of alternative transcription events, alternative RNA splicing events, and alternative translational and posttranslational events.

The invention includes an immunogen which includes an *H. pylori* polypeptide in an immunogenic preparation, the immunogen being capable of eliciting an immune response specific for said *H. pylori* polypeptide, e.g., a humoral response, an antibody response, or a cellular response. In preferred embodiments, the immunogen comprising an antigenic determinant from a protein contained in Figures 1A-558A.

In another aspect, the invention provides a substantially pure nucleic acid having a nucleotide sequence which encodes an *H. pylori* polypeptide. In preferred embodiments: the encoded polypeptide has biological activity the encoded polypeptide has an amino acid sequence at least 60%, 80%, 90%, 95%, 98%, or 99% homologous to an amino acid sequence contained in Figures 1A-558A; the encoded polypeptide has an amino acid sequence essentially the same as an amino acid sequence in Figures 1A-558A; the encoded polypeptide is at least 5, 10, 20, 50, 100, or 150 amino acids in length; the encoded polypeptide comprises at least 5, preferably at least 10, more preferably at least 20, more preferably at least 50, 100, or 150 contiguous amino acids contained in Figures 1A-558A.

In preferred embodiments: the nucleic acid is that as shown in Figure 559; the nucleic acid is at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% homologous with a nucleic acid sequence contained in Figure 559.

In a preferred embodiment, the encoded *H. pylori* polypeptide differs in amino acid sequence at 1, 2, 3, 5, 10 or more residues, from a sequence in Figures 1A-558A. The differences, however, are such that: the *H. pylori* encoded polypeptide exhibits a *H. pylori* biological activity, e.g., the encoded *H. pylori* enzyme retains a biological activity of a naturally occurring *H. pylori*.

In preferred embodiments, the encoded polypeptide includes all or a fragment of an amino acid sequence contained in Figures 1A-558A; fused, in reading frame, to additional

amino acid residues, preferably to residues encoded by genomic DNA 5' to the genomic DNA which encodes a sequence contained in Figures 1A-558A.

In preferred embodiments, the subject *H. pylori* nucleic acid will include a transcriptional regulatory sequence, e.g. at least one of a transcriptional promoter or transcriptional enhancer sequence, operably linked to the *H. pylori* gene sequence, e.g., to render the *H. pylori* gene sequence suitable for expression in a recombinant host cell.

In yet a further preferred embodiment, the nucleic acid which encodes an *H. pylori* polypeptide of the invention, hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 12 consecutive nucleotides contained in Figure 559; more preferably to at least 20 consecutive nucleotides contained in Figure 559; more preferably to at least 40 consecutive nucleotides contained in Figure 559.

In a preferred embodiment, the nucleic acid encodes a peptide which differs by at least one amino acid residue from the sequences shown in Figures 1A-558A.

In a preferred embodiment, the nucleic acid differs by at least one nucleotide from a nucleotide sequence shown in Figure 559 which encodes amino acids shown in Figures 1A-558A.

In another aspect, the invention includes: a vector including a nucleic acid which encodes an *H. pylori*-like polypeptide, e.g., an *H. pylori* polypeptide; a host cell transfected with the vector; and a method of producing a recombinant *H. pylori*-like polypeptide, e.g., an *H. pylori* polypeptide; including culturing the cell, e.g., in a cell culture medium, and isolating the *H. pylori*-like polypeptide, e.g., an *H. pylori* polypeptide, e.g., from the cell or from the cell culture medium.

In another aspect, the invention features, a purified recombinant nucleic acid having at least 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% homology with a sequence contained in Figure 559.

The invention also provides a probe or primer which includes a substantially purified oligonucleotide. The oligonucleotide includes a region of nucleotide sequence which hybridizes under stringent conditions to at least 10 consecutive nucleotides of sense or antisense sequence contained in Figure 559, or naturally occurring mutants thereof. In preferred embodiments, the probe or primer further includes a label group attached thereto. The label group can be, e.g., a radioisotope, a fluorescent compound, an enzyme, and/or an enzyme co-factor. Preferably the oligonucleotide is at least 10 and less than 20, 30, 50, 100, or 150 nucleotides in length.

The invention involves nucleic acids, e.g., RNA or DNA, encoding a polypeptide of the invention. This includes double stranded nucleic acids as well as coding and antisense single strands.

The *H. pylori* strain, from which genomic sequences have been sequenced, has been deposited in the American Type Culture Collection(ATCC) as strain HP-J99.

The nucleic acid sequences of this invention may be obtained directly from the DNA of the above referenced *H. pylori* strain by using the polymerase chain reaction (PCR). See "PCR, A Practical Approach" (McPherson, Quirke, and Taylor, eds., IRL Press, Oxford, UK, 1991) for details about the PCR. High fidelity PCR can be used to ensure a faithful DNA copy prior to expression. In addition, amplified products can be checked by conventional sequencing methods. Clones carrying the desired sequences described in this invention may be obtained by screening the libraries by means of the PCR or by hybridization of synthetic oligonucleotide probes to filter lifts of the library colonies or plaques as known in the art (see, eg, Sambrook et al., *Molecular Cloning, A Laboratory Manual* 2nd edition, 1989, Cold Spring Harbor Press, NY).

Nucleic acids isolated or synthesized in accordance with features of the present invention are useful, by way of example, without limitation, as probes, primers, capture ligands, antisense genes and for developing expression systems for the synthesis of proteins and peptides corresponding to such sequences.

As probes, primers, capture ligands and antisense agents, the nucleic acid will normally comprise approximately twenty or more nucleotides for specificity as well as the ability to form stable hybridization products.

Probes

A nucleic acid isolated or synthesized in accordance with Figure 559 can be used as a probe to specifically detect *H. pylori*. With the sequence information set forth in the present application, sequences of twenty or more nucleotides are identified which provide the desired inclusivity and exclusivity with respect to *H. pylori*, and extraneous nucleic acid sequences likely to be encountered during hybridization conditions. More preferably, the sequence will comprise at least twenty to thirty nucleotides to convey stability to the hybridization product formed between the probe and the intended target molecules.

Sequences larger than 1000 nucleotides in length are difficult to synthesize but can be generated by recombinant DNA techniques. Individuals skilled in the art will readily recognize that the nucleic acid sequences, for use as probes, can be provided with a label to facilitate detection of a hybridization product.

Nucleic acid isolated and synthesized in accordance with Figure 559 may also be useful as probes to detect homologous regions (especially homologous genes) of other *Helicobacter* species using relaxed stringency hybridization conditions, as will be obvious to anybody skilled in the art.

Capture Ligand

For use as a capture ligand, the nucleic acid selected in the manner described above with respect to probes, can be readily associated with a support. The manner in which nucleic acid is associated with supports is well known. Nucleic acid having twenty or more

nucleotides in a sequence contained in Figure 559 have utility to separate *H. pylori* nucleic acid from the nucleic acid of each other and other organisms. Nucleic acid having twenty or more nucleotides in a sequence shown in Figure 559 may also have utility to separate other *Helicobacter* species from each other and from other organisms. Preferably, the sequence will comprise at least twenty nucleotides to convey stability to the hybridization product formed between the probe and the intended target molecules. Sequences larger than 1000 nucleotides in length are difficult to synthesize but can be generated by recombinant DNA techniques.

Primers

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility as primers for the amplification of *H. pylori* nucleic acid. These nucleic acids may also have utility as primers for the amplification of nucleic acid sequences in other *Helicobacter* species. With respect to polymerase chain reaction (PCR) techniques, nucleic acid sequences of \geq 10-15 nucleotides contained in Figure 559 have utility in conjunction with suitable enzymes and reagents to create copies of *H. pylori* nucleic acid. More preferably, the sequence will comprise twenty or more nucleotides to convey stability to the hybridization product formed between the primer and the intended target molecules. Binding conditions of primers greater than 100 nucleotides are more difficult to control to obtain specificity. High fidelity PCR can be used to ensure a faithful DNA copy prior to expression. In addition, amplified products can be checked by conventional sequencing methods.

The copies can be used in diagnostic assays to detect specific sequences, including genes *H. pylori* and/or other *Helicobacter* species. The copies can also be incorporated into cloning and expression vectors to generate polypeptides corresponding to the nucleic acid synthesized by PCR, as will be described in greater detail below.

Antisense

Nucleic acid or nucleic acid-hybridizing derivatives isolated or synthesized in accordance with the sequences described herein have utility as antisense agents to prevent the expression of *H. pylori* genes. These sequences may also have utility as antisense agents to prevent expression of genes of other *Helicobacter* species.

Nucleic acid or derivatives corresponding to *H. pylori* nucleic acid sequences is loaded into a suitable carrier such as a liposome or bacteriophage for introduction into bacterial cells. For example, a nucleic acid having twenty or more nucleotides is capable of binding to bacteria nucleic acid or bacteria messenger RNA. Preferably, the antisense nucleic acid is comprised of 20 or more nucleotides to provide necessary stability of a hybridization product of non-naturally occurring nucleic acid and bacterial nucleic acid and/or bacterial messenger RNA. Nucleic acid having a sequence greater than 1000 nucleotides in length is difficult to synthesize but can be generated by recombinant DNA techniques. Methods for

loading antisense nucleic acid in liposomes is known in the art as exemplified by U.S. Patent 4,241,046 issued December 23, 1980 to Papahadjopoulos et al.

Expressing *H. pylori* Genes

The function of a specific gene or operon can be ascertained by expression in a bacterial strain under conditions where the activity of the gene product(s) specified by the gene or operon in question can be specifically measured. Alternatively, a gene product may be produced in large quantities in an expressing strain for use as an antigen, an industrial reagent, for structural studies, etc. This expression could be accomplished in a mutant strain which lacks the activity of the gene to be tested, or in a strain that does not produce the same gene product(s). This includes, but is not limited to other *Helicobacter* strains, and other bacterial strains such as *E. coli*, *Norcardia*, *Corynebacterium*, and *Streptomyces* species. In some cases the expression host will utilize the natural *Helicobacter* promoter whereas in others, it will be necessary to drive the gene with a promoter sequence derived from the expressing organism (e.g., an *E. coli* beta-galactosidase promoter for expression in *E. coli*).

To express a gene product using the natural *H. pylori* promoter, a procedure such as the following is used. A restriction fragment containing the gene of interest, together with its associated natural promoter element and regulatory sequences (identified using the DNA sequence data) is cloned into an appropriate recombinant plasmid containing the following components: an origin of replication that functions in the host organism, and an appropriate selectable marker. This can be accomplished by a number of procedures known to those skilled in the art. It is most preferably done by cutting the plasmid and the fragment to be cloned with the same restriction enzyme to produce compatible ends that can be ligated to join the two pieces together. The recombinant plasmid is introduced into the host organism by electroporation and cells containing the recombinant plasmid are identified by selection for the marker on the plasmid. Expression of the desired gene product is detected using an assay specific for that gene product.

In the case of a gene that requires a different promoter, the body of the gene (coding sequence) is specifically excised and cloned into an appropriate expression plasmid. This subcloning can be done by several methods, but is most easily accomplished by PCR amplification of a specific fragment and ligation into to expression plasmid after treating the PCR product with a restriction enzyme or exonuclease to create suitable ends for cloning.

Expressed Genes in Therapeutics

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility to generate proteins and peptides. The nucleic acid exemplified in Figure 559 can be cloned into suitable vectors or used to isolate nucleic acid. The isolated nucleic acid is combined with suitable DNA linkers and cloned into a suitable vector. The vector can

be used to transform a suitable host organism such as *E. coli* and the peptide or protein encoded by the sequences isolated.

Molecular cloning techniques are described in the text *Molecular Cloning: A Laboratory Manual*, 2nd edition, Sambrook et al., Coldspring Harbor Laboratory, (1989).

Drug Screening Assays

By making available purified and recombinant *H. pylori* polypeptides, the present invention provides assays which can be used to screen for drugs which are either agonists or antagonists of the normal cellular function, in this case, of the subject *H. pylori* polypeptides, or of their role in intracellular signaling. Such inhibitors or potentiators may be useful as new therapeutic agents to combat *H. pylori* infections in humans. A variety of assay formats will suffice and, in light of the present inventions, will be comprehended by skilled artisan.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the *in vitro* system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with other proteins or change in enzymatic properties of the molecular target. Accordingly, in an exemplary screening assay of the present invention, the compound of interest is contacted with an isolated and purified *H. pylori* polypeptide.

Screening assays may be constructed *in vitro* with purified *H. pylori* enzyme such that the action of the enzyme produces an easily detectable reaction product. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. Suitable products include those with distinctive absorption, fluorescence, or chemi-luminescence properties, for example, because detection may be easily automated. A variety of synthetic or naturally occurring compounds may be tested in the assay to identify those which inhibit or potentiate the activity of the *H. pylori* enzyme. Some of these active compounds may directly, or with chemical alterations to promote membrane permeability or solubility, also inhibit or potentiate the same enzymatic activity in whole, live *H. pylori* cells.

Antibodies

The invention also includes antibodies specifically reactive with the subject *H. pylori*-like polypeptide. Anti-protein/anti-peptide antisera or monoclonal antibodies can be made by

standard protocols (See, for example, *Antibodies: A Laboratory Manual* ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide. Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of the subject *H. pylori* polypeptide can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies. In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of the *H. pylori* polypeptides of the invention, e.g. antigenic determinants of a polypeptide shown in Figures 1A-*B, or a closely related human or non-human mammalian homolog (e.g. 90% percent homologous, more preferably at least 95 percent homologous). In yet a further preferred embodiment of the present invention, the anti-*H. pylori* antibodies do not substantially cross react (i.e. react specifically) with a protein which is: e.g., less than 80% percent homologous to a sequence shown in Figure 1A-558A. By "not substantially cross react", it is meant that the antibody has a binding affinity for a non-homologous protein which is less than 10 percent, more preferably less than 5 percent, and even more preferably less than 1 percent, of the binding affinity for a protein contained in Figure 1A-558A. In a most preferred embodiment, there is no crossreactivity between bacterial and mammalian antigens.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with *H. pylori* polypeptides. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, $F(ab')_2$ fragments can be generated by treating antibody with pepsin. The resulting $F(ab')_2$ fragment can be treated to reduce disulfide bridges to produce Fab' fragments. The antibody of the present invention is further intended to include bispecific and chimeric molecules having an anti-*H. pylori* portion.

Both monoclonal and polyclonal antibodies (Ab) directed against *H. pylori* polypeptides or *H. pylori* polypeptide variants, and antibody fragments such as Fab' and $F(ab')_2$, can be used to block the action of *H. pylori* polypeptide and allow the study of the role of a particular *H. pylori* polypeptide of the present invention in aberrant or unwanted intracellular signaling, as well as the normal cellular function of the *H. pylori* and by microinjection of anti-*H. pylori* polypeptide antibodies of the present invention.

Antibodies which specifically bind *H. pylori* epitopes can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of *H. pylori* antigens. Anti *H. pylori* polypeptide antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate *H. pylori* levels in tissue or bodily fluid as part of a clinical testing procedure. Likewise, the ability to monitor *H. pylori* polypeptide levels in an individual can allow determination of the efficacy

of a given treatment regimen for an individual afflicted with such a disorder. The level of an *H. pylori* polypeptide can be measured in cells found in bodily fluid, such as in urine samples or can be measured in tissue, such as produced by gastric biopsy. Diagnostic assays using anti-*H. pylori* antibodies can include, for example, immunoassays designed to aid in early diagnosis of *H. pylori* infections. The present invention can also be used as a method of detecting antibodies contained in samples from individuals infected by this bacterium using specific *H. pylori* antigens.

Another application of anti-*H. pylori* polypeptide antibodies of the present invention is in the immunological screening of cDNA libraries constructed in expression vectors such as λ gt11, λ gt18-23, λ ZAP, and λ ORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, λ gt11 will produce fusion proteins whose amino termini consist of β -galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a subject *H. pylori* polypeptide can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with anti-*H. pylori* polypeptide antibodies. Phage, scored by this assay, can then be isolated from the infected plate. Thus, the presence of *H. pylori* gene homologs can be detected and cloned from other species, and alternate isoforms (including splicing variants) can be detected and cloned.

Vaccines

The present invention also includes vaccine compositions for protection against infection by *H. pylori*, a gram-negative spiral microaerophilic bacterium. In one embodiment, the vaccine compositions contain immunogenic surface proteins from *H. pylori*, or portion thereof, and a pharmaceutically acceptable carrier. Nucleic acids within the scope of the invention are exemplified by the nucleic acids shown in Figure 559 and which encode *H. pylori* surface proteins shown in Figures 1A - 558A. However, any nucleic acid encoding an immunogenic *H. pylori* protein, or portion thereof, which is capable of expression in a cell, can be used in the present invention. These vaccines can have therapeutic and prophylactic utilities.

Another aspect of the present invention provides vaccine compositions for protection against infection by *H. pylori* which contain a modified immunogenic *H. pylori* protein or portion thereof, or a nucleic acid encoding modified *H. pylori* protein or portion thereof, and a pharmaceutically acceptable carrier. It is possible to modify the structure of a *H. pylori* protein or peptide for such purposes as increasing solubility, enhancing stability (e.g., shelf life *ex vivo* and resistance to proteolytic degradation *in vivo*). A modified *H. pylori* protein or peptide can be produced in which the amino acid sequence has been altered, such as by amino acid substitution, deletion, or addition.

Another example of modification of an *H. pylori* peptide is substitution of cysteine residues preferably with alanine, serine, threonine, leucine or glutamic acid residues to minimize dimerization via disulfide linkages. In addition, amino acid side chains of fragments of the protein of the invention can be chemically modified. Another modification is cyclization of the peptide.

In order to enhance stability and/or reactivity, an *H. pylori* protein or peptide can be modified to incorporate one or more polymorphisms in the amino acid sequence of the protein resulting from any natural allelic variation. Additionally, D-amino acids, non-natural amino acids, or non-amino acid analogs can be substituted or added to produce a modified protein within the scope of this invention. Furthermore, an *H. pylori* protein can be modified using polyethylene glycol (PEG) according to the method of A. Sehon and co-workers (Wie et al., *supra*) to produce a protein conjugated with PEG. In addition, PEG can be added during chemical synthesis of the protein. Other modifications of *H. pylori* proteins include reduction/alkylation (Tarr, *Methods of Protein Microcharacterization*, J. E. Silver ed., Humana Press, Clifton NJ 155-194 (1986)); acylation (Tarr, *supra*); chemical coupling to an appropriate carrier (Mishell and Shiigi, eds, *Selected Methods in Cellular Immunology*, WH Freeman, San Francisco, CA (1980), U.S. Patent 4,939,239; or mild formalin treatment (Marsh, (1971) *Int. Arch. of Allergy and Appl. Immunol.*, 41: 199 - 215).

To facilitate purification and potentially increase solubility of an *H. pylori* protein or peptide, it is possible to add an amino acid fusion moiety to the peptide backbone. For example, hexa-histidine can be added to the protein for purification by immobilized metal ion affinity chromatography (Hochuli, E. et al., (1988) *Bio/Technology*, 6: 1321 - 1325). In addition, to facilitate isolation of peptides free of irrelevant sequences, specific endoprotease cleavage sites can be introduced between the sequences of the fusion moiety and the peptide.

To potentially aid proper antigen processing of T cell epitopes within an *H. pylori* protein of the invention, canonical protease sensitive sites can be engineered between regions, each comprising at least one T cell epitope via recombinant or synthetic methods. For example, charged amino acid pairs, such as KK or RR, can be introduced between regions within a protein or fragment during recombinant construction thereof. The resulting peptide can be rendered sensitive to cleavage by cathepsin and/or other trypsin-like enzymes which would generate portions of the protein containing one or more T cell epitopes. In addition, such charged amino acid residues can result in an increase in the solubility of the peptide.

Site-directed mutagenesis of a nucleic acid encoding an *H. pylori* protein can be used to modify the structure of the peptide by methods known in the art. Such methods may, among others, include polymerase chain reaction (PCR) with oligonucleotide primers bearing one or more mutations (Ho et al., (1989) *Gene*, 77: 51 - 59) or total synthesis of mutated genes (Hostomsky, Z. et al., (1989) *Biochem. Biophys. Res. Comm.*, 161: 1056 - 1063). To enhance recombinant protein expression, the aforementioned methods can be applied to

change the codons present in the cDNA sequence of the invention to those preferentially utilized by the host cell in which the recombinant protein is being expressed (Wada et al., supra). An extensive discussion of mutagenesis protocols is provided in the "Production of Fragments and Analogs" section herein.

Another aspect of the invention provides a vaccine composition for protection against infection by *H. pylori* which contains an immunogenic fragment of an *H. pylori* protein or portion thereof, and a pharmaceutically acceptable carrier. Preferred fragments include peptides of at least about 10 amino acid residues in length, preferably about 10-20 amino acid residues in length, and more preferably about 12-16 amino acid residues in length.

Immunogenic peptides of the invention can be obtained, for example, by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid encoding the full-length *H. pylori* protein. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, an *H. pylori* protein may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptides having the ability to induce a T cell response, such as stimulation (proliferation, cytokine secretion). An extensive discussion of peptide analogs and fragments is provided in the "Production of Fragments and Analogs" section herein.

In one embodiment, immunogenic *H. pylori* fragments can be identified by the ability of the peptide to stimulate T cells. Peptides which stimulate T cells, as determined by, for example, T cell proliferation or cytokine secretion are defined herein as comprising at least one T cell epitope. T cell epitopes are believed to be involved in initiation and perpetuation of the immune response to the protein allergen which is responsible for the clinical symptoms of allergy. These T cell epitopes are thought to trigger early events at the level of the T helper cell by binding to an appropriate HLA molecule on the surface of an antigen presenting cell, thereby stimulating the T cell subpopulation with the relevant T cell receptor for the epitope. These events lead to T cell proliferation, lymphokine secretion, local inflammatory reactions, recruitment of additional immune cells to the site of antigen/T cell interaction, and activation of the B cell cascade, leading to the production of antibodies. One isotype of these antibodies, IgE, is fundamentally important to the development of allergic symptoms and its production is influenced early in the cascade of events at the level of the T helper cell, by the nature of the lymphokines secreted. A T cell epitope is the basic element, or smallest unit of recognition by a T cell receptor, where the epitope comprises amino acids essential to receptor recognition. Amino acid sequences which mimic those of the T cell epitopes and which modify the allergic response to protein allergens are within the scope of this invention.

Screening peptides for those which are immunogenic can be accomplished using one or more of several different assays. For example, *in vitro*, peptide T cell stimulatory activity is assayed by contacting a peptide known or suspected of being immunogenic with an antigen presenting cell which presents appropriate MHC molecules in a T cell culture. Presentation of an immunogenic *H. pylori* peptide in association with appropriate MHC molecules to T cells in conjunction with the necessary costimulation has the effect of transmitting a signal to the T cell that induces the production of increased levels of cytokines, particularly of interleukin-2 and interleukin-4. The culture supernatant can be obtained and assayed for interleukin-2 or other known cytokines. For example, any one of several conventional assays for interleukin-2 can be employed, such as the assay described in *Proc. Natl. Acad. Sci USA*, 86: 1333 (1989) the pertinent portions of which are incorporated herein by reference. A kit for an assay for the production of interferon is also available from Genzyme Corporation (Cambridge, MA).

Alternatively, a common assay for T cell proliferation entails measuring tritiated thymidine incorporation. The proliferation of T cells can be measured *in vitro* by determining the amount of ^3H -labeled thymidine incorporated into the replicating DNA of cultured cells. Therefore, the rate of DNA synthesis and, in turn, the rate of cell division can be quantified.

To produce modified proteins or immunogenic fragments by recombinant DNA techniques, an expression vector containing a nucleic acid encoding all or a portion of a *H. pylori* protein, operably linked to at least one regulatory sequence can be used. Operably linked is intended to mean that the nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleotide sequence. Regulatory sequences are art-recognized and include promoters, enhancers and other expression control elements. Such regulatory sequences are described in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed. In one embodiment of the present invention, the expression vector includes nucleic acid, preferably a DNA, encoding a modified *H. pylori* protein or immunogenic fragment having all or a portion of the amino acid sequence. Such expression vectors can be used to transfect cells to thereby produce proteins or peptides, including fusion proteins or peptides encoded by nucleic acids as described herein.

Host cells suitable for transfection and recombinant production of *H. pylori* proteins of the invention include any prokaryotic or eucaryotic cell. For example, an *H. pylori* protein or peptide may be expressed in bacterial cells such as *E. coli*, insect cells (baculovirus), yeast, or mammalian cells such as Chinese hamster ovary cells (CHO). Other suitable host cells can be found in Goeddel, (1990) *supra* or known to those skilled in the art.

H. pylori proteins and fragments of the invention can also be chemically synthesized, using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. The nucleic acids of the invention can also be chemically synthesized using standard techniques. Various methods of chemically synthesizing polydeoxynucleotides are known, including solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (see e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al., U.S. Patent No. 4,458,066; and Itakura, U.S. Patent Nos. 4,401,796 and 4,373,071, incorporated by reference herein).

Vaccine compositions of the present invention containing DNA encoding immunogenic protein from *H. pylori*, or containing modified protein or fragments, contain both the DNA or protein and a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier that does not cause an allergic reaction or other untoward effect in patients to whom it is administered.

Suitable pharmaceutically acceptable carriers include, for example, one or more of water, saline, phosphate buffered saline, dextrose, glycerol, ethanol and the like, as well as combinations thereof. Pharmaceutically acceptable carriers may further comprise minor amounts of auxiliary substances such as wetting or emulsifying agents, preservatives or buffers, which enhance the shelf life or effectiveness of the antibody. For vaccines of the invention containing modified *H. pylori* protein or immunogenic protein fragments, the protein or peptide must be coadministered with a suitable adjuvant.

It will be apparent to those of skill in the art that the therapeutically effective amount of DNA or protein of this invention will depend, *inter alia*, upon the administration schedule, the unit dose of antibody administered, whether the protein or DNA is administered in combination with other therapeutic agents, the immune status and health of the patient, and the therapeutic activity of the particular protein or DNA.

Vaccine compositions are conventionally administered parenterally, e.g., by injection, either subcutaneously or intramuscularly. Methods for intramuscular immunization are described by Wolff et al. (1990) *Science* 247: 1465-1468 and by Sedegah et al. (1994) *Immunology* 91: 9866-9870. Other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Oral immunization is preferred over parenteral methods for inducing protection against infection by *H. pylori*. Czinn et. al. (1993) *Vaccine* 11: 637-642. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like.

The selection of candidate protein antigens for vaccine development were derived from the nucleotide sequence. First, all possible open reading frames (ORF) greater than 50 nucleotides in all six reading frames were identified and translated into amino acid sequences. Second, the identified ORF's were analyzed for homology to other known exported or membrane proteins and the ORF's were also analyzed using the discriminant analysis

described by Klein, et al. (Klein, P., Kanehsia, M., and DeLisi, C. (1985) *Biochimica et Biophysica Acta* 815, 468-476) for predicting exported and membrane proteins.

Homology searches were performed using the BLAST algorithm contained in the Wisconsin Sequence Analysis Package (Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) to compare each predicted ORF amino acid sequence with all sequences found in the current GenBank, SWISS-PROT and PIR databases. BLAST searches for local alignments between the ORF and the databank sequences and reports a probability score which indicates the probability of finding this sequence by chance in the database. ORF's with significant homology (e.g. probabilities better than 1×10^{-6} (ee-6)) to membrane or exported proteins represent likely protein antigens for vaccine development. Possible functions are provided to some of the *H. pylori* genes as indicated in Figure 560 based on sequence homology to genes cloned in other organisms.

Discriminant analysis (Klein, et al. *supra*) was used to examine the ORF amino acid sequences using our own software. This algorithm uses the intrinsic information contained in the ORF amino acid sequence and compares it to information derived from the properties of known membrane and exported proteins. This comparison predicts which proteins will be exported, membrane associated or cytoplasmic. ORF amino acid sequences identified as exported or membrane associated by this algorithm are likely protein antigens for vaccine development.

Infrequently it is not possible to distinguish between multiple possible nucleotides at a given position in the nucleic acid sequence. In those cases the ambiguities are denoted by an extended alphabet as follows:

These are the official IUPAC-IUB single-letter base codes

Code	Base Description
G	Guanine
A	Adenine
T	Thymine
C	Cytosine
R	Purine (A or G)
Y	Pyrimidine (C or T or U)
M	Amino (A or C)
K	Ketone (G or T)
S	Strong interaction (C or G)
W	Weak interaction (A or T)
H	Not-G (A or C or T)
B	Not-A (C or G or T)
V	Not-T (not-U) (A or C or G)
D	Not-C (A or G or T)
N	Any (A or C or G or T)

The amino acid translations of this invention account for the ambiguity in the nucleic acid sequence by translating the ambiguous codon as the letter "X". In all cases, the

permissible amino acid residues at a position are clear from an examination of the nucleic acid sequence based on the standard genetic code.

Production of Fragments and Analogs

The inventor has discovered novel gene products, e.g. bacterial surface gene products, from the organism *H. pylori*. Once an example of this core structure has been provided one skilled in the art can alter the disclosed structure (of *H. pylori* genes), e.g., by producing fragments or analogs, and test the newly produced structures for activity. Examples of prior art methods which allow the production and testing of fragments and analogs are discussed below. These, or analogous methods can be used to make and screen libraries of polypeptides, e.g., libraries of random peptides or libraries of fragments or analogues of cellular proteins for the ability to bind *H. pylori* polypeptides. Such screens are useful for discovery of inhibitors of *H. pylori*.

Generation of Fragments

Fragments of a protein can be produced in several ways, e.g., recombinantly, by proteolytic digestion, or by chemical synthesis. Internal or terminal fragments of a polypeptide can be generated by removing one or more nucleotides from one end (for a terminal fragment) or both ends (for an internal fragment) of a nucleic acid which encodes the polypeptide. Expression of the mutagenized DNA produces polypeptide fragments. Digestion with "end-nibbling" endonucleases can thus generate DNA's which encode an array of fragments. DNA's which encode fragments of a protein can also be generated by random shearing, restriction digestion or a combination of the above-discussed methods.

Fragments can also be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, peptides of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or divided into overlapping fragments of a desired length.

Production of Altered DNA and Peptide Sequences: Random Methods

Amino acid sequence variants of a protein can be prepared by random mutagenesis of DNA which encodes a protein or a particular domain or region of a protein. Useful methods include PCR mutagenesis and saturation mutagenesis. A library of random amino acid sequence variants can also be generated by the synthesis of a set of degenerate oligonucleotide sequences. (Methods for screening proteins in a library of variants are elsewhere herein).

PCR Mutagenesis

In PCR mutagenesis, reduced Taq polymerase fidelity is used to introduce random mutations into a cloned fragment of DNA (Leung et al., 1989, *Technique* 1:11-15). This is a

very powerful and relatively rapid method of introducing random mutations. The DNA region to be mutagenized is amplified using the polymerase chain reaction (PCR) under conditions that reduce the fidelity of DNA synthesis by Taq DNA polymerase, e.g., by using a dGTP/dATP ratio of five and adding Mn^{2+} to the PCR reaction. The pool of amplified DNA fragments are inserted into appropriate cloning vectors to provide random mutant libraries.

Saturation Mutagenesis

Saturation mutagenesis allows for the rapid introduction of a large number of single base substitutions into cloned DNA fragments (Mayers et al., 1985, *Science* 229:242). This technique includes generation of mutations, e.g., by chemical treatment or irradiation of single-stranded DNA *in vitro*, and synthesis of a complimentary DNA strand. The mutation frequency can be modulated by modulating the severity of the treatment, and essentially all possible base substitutions can be obtained. Because this procedure does not involve a genetic selection for mutant fragments both neutral substitutions, as well as those that alter function, are obtained. The distribution of point mutations is not biased toward conserved sequence elements.

Degenerate Oligonucleotides

A library of homologs can also be generated from a set of degenerate oligonucleotide sequences. Chemical synthesis of a degenerate sequences can be carried out in an automatic DNA synthesizer, and the synthetic genes then ligated into an appropriate expression vector. The synthesis of degenerate oligonucleotides is known in the art (see for example, Narang, SA (1983) *Tetrahedron* 39:3; Itakura et al. (1981) *Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules*, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) *Science* 249:386-390; Roberts et al. (1992) *PNAS* 89:2429-2433; Devlin et al. (1990) *Science* 249: 404-406; Cwirla et al. (1990) *PNAS* 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

Production of Altered DNA and Peptide Sequences: Methods for Directed Mutagenesis

Non-random or directed, mutagenesis techniques can be used to provide specific sequences or mutations in specific regions. These techniques can be used to create variants which include, e.g., deletions, insertions, or substitutions, of residues of the known amino acid sequence of a protein. The sites for mutation can be modified individually or in series, e.g., by (1) substituting first with conserved amino acids and then with more radical choices

depending upon results achieved, (2) deleting the target residue, or (3) inserting residues of the same or a different class adjacent to the located site, or combinations of options 1-3.

Alanine Scanning Mutagenesis

Alanine scanning mutagenesis is a useful method for identification of certain residues or regions of the desired protein that are preferred locations or domains for mutagenesis, Cunningham and Wells (*Science* 244:1081-1085, 1989). In alanine scanning, a residue or group of target residues are identified (e.g., charged residues such as Arg, Asp, His, Lys, and Glu) and replaced by a neutral or negatively charged amino acid (most preferably alanine or polyalanine). Replacement of an amino acid can affect the interaction of the amino acids with the surrounding aqueous environment in or outside the cell. Those domains demonstrating functional sensitivity to the substitutions are then refined by introducing further or other variants at or for the sites of substitution. Thus, while the site for introducing an amino acid sequence variation is predetermined, the nature of the mutation per se need not be predetermined. For example, to optimize the performance of a mutation at a given site, alanine scanning or random mutagenesis may be conducted at the target codon or region and the expressed desired protein subunit variants are screened for the optimal combination of desired activity.

Oligonucleotide-Mediated Mutagenesis

Oligonucleotide-mediated mutagenesis is a useful method for preparing substitution, deletion, and insertion variants of DNA, see, e.g., Adelman et al., (*DNA* 2:183, 1983). Briefly, the desired DNA is altered by hybridizing an oligonucleotide encoding a mutation to a DNA template, where the template is the single-stranded form of a plasmid or bacteriophage containing the unaltered or native DNA sequence of the desired protein. After hybridization, a DNA polymerase is used to synthesize an entire second complementary strand of the template that will thus incorporate the oligonucleotide primer, and will code for the selected alteration in the desired protein DNA. Generally, oligonucleotides of at least 25 nucleotides in length are used. An optimal oligonucleotide will have 12 to 15 nucleotides that are completely complementary to the template on either side of the nucleotide(s) coding for the mutation. This ensures that the oligonucleotide will hybridize properly to the single-stranded DNA template molecule. The oligonucleotides are readily synthesized using techniques known in the art such as that described by Crea et al. (*Proc. Natl. Acad. Sci. USA*, 75: 5765[1978]).

Cassette Mutagenesis

Another method for preparing variants, cassette mutagenesis, is based on the technique described by Wells et al. (*Gene*, 34:315[1985]). The starting material is a plasmid (or other vector) which includes the protein subunit DNA to be mutated. The codon(s) in the

protein subunit DNA to be mutated are identified. There must be a unique restriction endonuclease site on each side of the identified mutation site(s). If no such restriction sites exist, they may be generated using the above-described oligonucleotide-mediated mutagenesis method to introduce them at appropriate locations in the desired protein subunit DNA. After the restriction sites have been introduced into the plasmid, the plasmid is cut at these sites to linearize it. A double-stranded oligonucleotide encoding the sequence of the DNA between the restriction sites but containing the desired mutation(s) is synthesized using standard procedures. The two strands are synthesized separately and then hybridized together using standard techniques. This double-stranded oligonucleotide is referred to as the cassette. This cassette is designed to have 3' and 5' ends that are comparable with the ends of the linearized plasmid, such that it can be directly ligated to the plasmid. This plasmid now contains the mutated desired protein subunit DNA sequence.

Combinatorial Mutagenesis

Combinatorial mutagenesis can also be used to generate mutants (Ladner et al., WO 88/06630). In this method, the amino acid sequences for a group of homologs or other related proteins are aligned, preferably to promote the highest homology possible. All of the amino acids which appear at a given position of the aligned sequences can be selected to create a degenerate set of combinatorial sequences. The variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level, and is encoded by a variegated gene library. For example, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential sequences are expressible as individual peptides, or alternatively, as a set of larger fusion proteins containing the set of degenerate sequences.

Primary High-Through-Put Methods for Screening Libraries of Peptide Fragments or Homologs

Various techniques are known in the art for screening generated mutant gene products. Techniques for screening large gene libraries often include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the genes under conditions in which detection of a desired activity, e.g., in this case, binding to *H. pylori* polypeptide or an interacting protein, facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the techniques described below is amenable to high through-put analysis for screening large numbers of sequences created, e.g., by random mutagenesis techniques.

Two Hybrid Systems

Two hybrid assays such as the system described above (as with the other screening methods described herein), can be used to identify polypeptides, e.g., fragments or analogs of

a naturally-occurring *H. pylori* polypeptide, e.g., of cellular proteins, or of randomly generated polypeptides which bind to an *H. pylori* protein. (The *H. pylori* domain is used as the bait protein and the library of variants are expressed as fish fusion proteins.) In an analogous fashion, a two hybrid assay (as with the other screening methods described herein), can be used to find polypeptides which bind a *H. pylori* polypeptide.

Display Libraries

In one approach to screening assays, the candidate peptides are displayed on the surface of a cell or viral particle, and the ability of particular cells or viral particles to bind an appropriate receptor protein via the displayed product is detected in a "panning assay". For example, the gene library can be cloned into the gene for a surface membrane protein of a bacterial cell, and the resulting fusion protein detected by panning (Ladner et al., WO 88/06630; Fuchs et al. (1991) *Bio/Technology* 9:1370-1371; and Goward et al. (1992) *TIBS* 18:136-140). In a similar fashion, a detectably labeled ligand can be used to score for potentially functional peptide homologs. Fluorescently labeled ligands, e.g., receptors, can be used to detect homolog which retain ligand-binding activity. The use of fluorescently labeled ligands, allows cells to be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, to be separated by a fluorescence-activated cell sorter.

A gene library can be expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at concentrations well over 10^{13} phage per milliliter, a large number of phage can be screened at one time. Second, since each infectious phage displays a gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd., and f1 are most often used in phage display libraries. Either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle. Foreign epitopes can be expressed at the NH₂-terminal end of pIII and phage bearing such epitopes recovered from a large excess of phage lacking this epitope (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) *J. Biol. Chem.* 267:16007-16010; Griffiths et al. (1993) *EMBO J* 12:725-734; Clackson et al. (1991) *Nature* 352:624-628; and Barbas et al. (1992) *PNAS* 89:4457-4461).

A common approach uses the maltose receptor of *E. coli* (the outer membrane protein, LamB) as a peptide fusion partner (Charbit et al. (1986) *EMBO* 5, 3029-3037). Oligonucleotides have been inserted into plasmids encoding the LamB gene to produce peptides fused into one of the extracellular loops of the protein. These peptides are available

for binding to ligands, e.g., to antibodies, and can elicit an immune response when the cells are administered to animals. Other cell surface proteins, e.g., OmpA (Schorr et al. (1991) *Vaccines* 91, pp. 387-392), PhoE (Agterberg, et al. (1990) *Gene* 88, 37-45), and PAL (Fuchs et al. (1991) *Bio/Tech* 9, 1369-1372), as well as large bacterial surface structures have served as vehicles for peptide display. Peptides can be fused to pilin, a protein which polymerizes to form the pilus-a conduit for interbacterial exchange of genetic information (Thiry et al. (1989) *Appl. Environ. Microbiol.* 55, 984-993). Because of its role in interacting with other cells, the pilus provides a useful support for the presentation of peptides to the extracellular environment. Another large surface structure used for peptide display is the bacterial motive organ, the flagellum. Fusion of peptides to the subunit protein flagellin offers a dense array of many peptides copies on the host cells (Kuwajima et al. (1988) *Bio/Tech.* 6, 1080-1083). Surface proteins of other bacterial species have also served as peptide fusion partners. Examples include the *Staphylococcus* protein A and the outer membrane protease IgA of *Neisseria* (Hansson et al. (1992) *J. Bacteriol.* 174, 4239-4245 and Klauser et al. (1990) *EMBO J.* 9, 1991-1999).

In the filamentous phage systems and the LamB system described above, the physical link between the peptide and its encoding DNA occurs by the containment of the DNA within a particle (cell or phage) that carries the peptide on its surface. Capturing the peptide captures the particle and the DNA within. An alternative scheme uses the DNA-binding protein LacI to form a link between peptide and DNA (Cull et al. (1992) *PNAS USA* 89:1865-1869). This system uses a plasmid containing the LacI gene with an oligonucleotide cloning site at its 3'-end. Under the controlled induction by arabinose, a LacI-peptide fusion protein is produced. This fusion retains the natural ability of LacI to bind to a short DNA sequence known as LacO operator (LacO). By installing two copies of LacO on the expression plasmid, the LacI-peptide fusion binds tightly to the plasmid that encoded it. Because the plasmids in each cell contain only a single oligonucleotide sequence and each cell expresses only a single peptide sequence, the peptides become specifically and stably associated with the DNA sequence that directed its synthesis. The cells of the library are gently lysed and the peptide-DNA complexes are exposed to a matrix of immobilized receptor to recover the complexes containing active peptides. The associated plasmid DNA is then reintroduced into cells for amplification and DNA sequencing to determine the identity of the peptide ligands. As a demonstration of the practical utility of the method, a large random library of dodecapeptides was made and selected on a monoclonal antibody raised against the opioid peptide dynorphin B. A cohort of peptides was recovered, all related by a consensus sequence corresponding to a six-residue portion of dynorphin B. (Cull et al. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:1869)

This scheme, sometimes referred to as peptides-on-plasmids, differs in two important ways from the phage display methods. First, the peptides are attached to the C-terminus of the fusion protein, resulting in the display of the library members as peptides having free

carboxy termini. Both of the filamentous phage coat proteins, pIII and pVIII, are anchored to the phage through their C-termini, and the guest peptides are placed into the outward-extending N-terminal domains. In some designs, the phage-displayed peptides are presented right at the amino terminus of the fusion protein. (Cwirla, et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 6378-6382) A second difference is the set of biological biases affecting the population of peptides actually present in the libraries. The LacI fusion molecules are confined to the cytoplasm of the host cells. The phage coat fusions are exposed briefly to the cytoplasm during translation but are rapidly secreted through the inner membrane into the periplasmic compartment, remaining anchored in the membrane by their C-terminal hydrophobic domains, with the N-termini, containing the peptides, protruding into the periplasm while awaiting assembly into phage particles. The peptides in the LacI and phage libraries may differ significantly as a result of their exposure to different proteolytic activities. The phage coat proteins require transport across the inner membrane and signal peptidase processing as a prelude to incorporation into phage. Certain peptides exert a deleterious effect on these processes and are underrepresented in the libraries (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251). These particular biases are not a factor in the LacI display system.

The number of small peptides available in recombinant random libraries is enormous. Libraries of 10^7 - 10^9 independent clones are routinely prepared. Libraries as large as 10^{11} recombinants have been created, but this size approaches the practical limit for clone libraries. This limitation in library size occurs at the step of transforming the DNA containing randomized segments into the host bacterial cells. To circumvent this limitation, an *in vitro* system based on the display of nascent peptides in polysome complexes has recently been developed. This display library method has the potential of producing libraries 3-6 orders of magnitude larger than the currently available phage/phagemid or plasmid libraries. Furthermore, the construction of the libraries, expression of the peptides, and screening, is done in an entirely cell-free format.

In one application of this method (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251), a molecular DNA library encoding 10^{12} decapeptides was constructed and the library expressed in an *E. coli* S30 *in vitro* coupled transcription/translation system. Conditions were chosen to stall the ribosomes on the mRNA, causing the accumulation of a substantial proportion of the RNA in polysomes and yielding complexes containing nascent peptides still linked to their encoding RNA. The polysomes are sufficiently robust to be affinity purified on immobilized receptors in much the same way as the more conventional recombinant peptide display libraries are screened. RNA from the bound complexes is recovered, converted to cDNA, and amplified by PCR to produce a template for the next round of synthesis and screening. The polysome display method can be coupled to the phage display system. Following several rounds of screening, cDNA from the enriched pool of polysomes was cloned into a phagemid vector. This vector serves as both a peptide expression vector,

displaying peptides fused to the coat proteins, and as a DNA sequencing vector for peptide identification. By expressing the polysome-derived peptides on phage, one can either continue the affinity selection procedure in this format or assay the peptides on individual clones for binding activity in a phage ELISA, or for binding specificity in a completion phage ELISA (Barret, et al. (1992) *Anal. Biochem* 204,357-364). To identify the sequences of the active peptides one sequences the DNA produced by the phagemid host.

Secondary Screens

The high through-put assays described above can be followed by secondary screens in order to identify further biological activities which will, e.g., allow one skilled in the art to differentiate agonists from antagonists. The type of a secondary screen used will depend on the desired activity that needs to be tested. For example, an assay can be developed in which the ability to inhibit an interaction between a protein of interest and its respective ligand can be used to identify antagonists from a group of peptide fragments isolated through one of the primary screens described above.

Therefore, methods for generating fragments and analogs and testing them for activity are known in the art. Once the core sequence of interest is identified, it is routine to perform for one skilled in the art to obtain analogs and fragments.

Peptide Mimetics

The invention also provides for reduction of the protein binding domains of the subject *H. pylori*-like family polypeptides, e.g., an *H. pylori* polypeptide, to generate mimetics, e.g. peptide or non-peptide agents. The peptide mimetics are able to disrupt binding of a *H. pylori* to its counter ligand, e.g., in the case of an *H. pylori* polypeptide binding to a naturally occurring ligand. The critical residues of a subject *H. pylori* polypeptide which are involved in molecular recognition of a polypeptide can be determined and used to generate *H. pylori*-derived peptidomimetics which competitively or noncompetatively inhibit binding of the *H. pylori* polypeptide with an interacting polypeptide (see, for example, "Peptide inhibitors of human papillomavirus protein binding to retinoblastoma gene protein" European patent applications EP-412,762A and EP-B31,080A). For example, scanning mutagenesis can be used to map the amino acid residues of a particular *H. pylori* polypeptide involved in binding an interacting polypeptide, peptidomimetic compounds (e.g. diazepine or isoquinoline derivatives) can be generated which mimic those residues in binding to an interacting polypeptide, and which therefore can inhibit binding of an *H. pylori* polypeptide to an interacting polypeptide and thereby interfere with the function of *H. pylori* polypeptide. For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed.,

ESCOM Publisher: Leiden, Netherlands, 1988), substituted gamma lactam rings (Garvey et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) *J Med Chem* 29:295; and Ewenson et al. in *Peptides: Structure and Function* (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), β -turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc Perkin Trans* 1:1231), and β -aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun* 126:419; and Dann et al. (1986) *Biochem Biophys Res Commun* 134:71).

Kits

The nucleic acid, peptides and antibodies of the present invention can be combined with other reagents and articles to form kits. Kits for diagnostic purposes typically comprise the nucleic acid, peptides or antibodies in vials or other suitable vessels. Kits typically comprise other reagents for performing hybridization reactions, polymerase chain reactions (PCR), or for reconstitution of lyophilized components, such as aqueous media, salts, buffers, and the like. Kits may also comprise reagents for sample processing such as detergents, chaotropic salts and the like. Kits may also comprise immobilization means such as particles, supports, wells, dipsticks and the like. Kits may also comprise labeling means such as dyes, developing reagents, radioisotopes, fluorescent agents, luminescent or chemiluminescent agents, enzymes, intercalating agents and the like. With the nucleic acid and amino acid sequence information provided herein, individuals skilled in art can readily assemble kits to serve their particular purpose.

Other Embodiments

Included in the invention are: allelic variations; natural mutants; induced mutants; proteins encoded by DNA that hybridizes under high or low stringency conditions to a nucleic acid which encodes a polypeptide as shown in Figure 559 (for definitions of high and low stringency see Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989, 6.3.1 - 6.3.6, hereby incorporated by reference); and, polypeptides specifically bound by antisera to *H. pylori* polypeptides, especially by antisera to an active site or binding domain of *H. pylori* polypeptide

The invention also includes fragments, preferably biologically active fragments, or analogs of *H. pylori* polypeptides. A biologically active fragment or analog is one having any *in vivo* or *in vitro* activity which is characteristic of the *H. pylori* polypeptides shown in Figures 1A-558A, or of other naturally occurring *H. pylori* polypeptides, e.g., one or more of the biological activities described above. Especially preferred are fragments which exist *in vivo*, e.g., fragments which arise from post transcriptional processing or which arise from translation of alternatively spliced RNA's. Fragments include those expressed in native or endogenous cells as well as those made in expression systems, e.g., in CHO cells. Because

peptides such as *H. pylori* polypeptides often exhibit a range of physiological properties and because such properties may be attributable to different portions of the molecule, a useful *H. pylori* fragment or *H. pylori* analog is one which exhibits a biological activity in any biological assay for *H. pylori* activity. Most preferably the fragment or analog possesses 10%, preferably 40%, or at least 90% of the activity of *H. pylori*, in any *in vivo* or *in vitro* assay.

Analogs can differ from naturally occurring *H. pylori* polypeptides in amino acid sequence or in ways that do not involve sequence, or both. Non-sequence modifications include changes in acetylation, methylation, phosphorylation, carboxylation, or glycosylation.

Preferred analogs include *H. pylori* polypeptides (or biologically active fragments thereof) whose sequences differ from the wild-type sequence by one or more conservative amino acid substitutions or by one or more non-conservative amino acid substitutions, deletions, or insertions which do not abolish the biological activity of the *H. pylori* polypeptide. Conservative substitutions typically include the substitution of one amino acid for another with similar characteristics, e.g., substitutions within the following groups: valine, glycine; glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Other conservative substitutions can be taken from the table below.

TABLE 3
CONSERVATIVE AMINO ACID REPLACEMENTS

For Amino Acid	Code	Replace with any of
Alanine	A	D-Ala, Gly, beta-Ala, L-Cys, D-Cys
Arginine	R	D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile, D-Met, D-Ile, Orn, D-Orn
Asparagine	N	D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln
Aspartic Acid	D	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln
Cysteine	C	D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr
Glutamine	Q	D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp

Glutamic Acid	E	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Glycine	G	Ala, D-Ala, Pro, D-Pro, β -Ala Acp
Isoleucine	I	D-Ile, Val, D-Val, Leu, D-Leu, Met, D-Met
Leucine	L	D-Leu, Val, D-Val, Leu, D-Leu, Met, D-Met
Lysine	K	D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn, D-Orn
Methionine	M	D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val
Phenylalanine	F	D-Phe, Tyr, D-Thr, L-Dopa, His, D-His, Trp, D-Trp, Trans-3,4, or 5-phenylproline, cis-3,4, or 5-phenylproline
Proline	P	D-Pro, L-1-thioazolidine-4-carboxylic acid, D-or L-1-oxazolidine-4-carboxylic acid
Serine	S	D-Ser, Thr, D-Thr, allo-Thr, Met, D-Met, Met(O), D-Met(O), L-Cys, D-Cys
Threonine	T	D-Thr, Ser, D-Ser, allo-Thr, Met, D-Met, Met(O), D-Met(O), Val, D-Val
Tyrosine	Y	D-Tyr, Phe, D-Phe, L-Dopa, His, D-His
Valine	V	D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met

Other analogs within the invention are those with modifications which increase peptide stability; such analogs may contain, for example, one or more non-peptide bonds (which replace the peptide bonds) in the peptide sequence. Also included are: analogs that include residues other than naturally occurring L-amino acids, e.g., D-amino acids or non-naturally occurring or synthetic amino acids, e.g., β or γ amino acids; and cyclic analogs.

As used herein, the term "fragment", as applied to a *H. pylori* analog, will ordinarily be at least about 20 residues, more typically at least about 40 residues, preferably at least about 60 residues in length. Fragments of *H. pylori* polypeptides can be generated by methods known to those skilled in the art. The ability of a candidate fragment to exhibit a

biological activity of *H. pylori* polypeptide can be assessed by methods known to those skilled in the art as described herein. Also included are *H. pylori* polypeptides containing residues that are not required for biological activity of the peptide or that result from alternative mRNA splicing or alternative protein processing events.

In order to obtain an *H. pylori* polypeptide, *H. pylori* polypeptide-encoding DNA can be introduced into an expression vector, the vector introduced into a cell suitable for expression of the desired protein, and the peptide recovered and purified, by prior art methods. Antibodies to the peptides and proteins can be made by immunizing an animal, e.g., a rabbit or mouse, and recovering anti-*H. pylori* polypeptide antibodies by prior art methods.

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments and methods described herein. Such equivalents are intended to be encompassed by the scope of the following claims.

Other embodiments are within the following claims.